RESEARCH DEPARTMENT

Ionospheric non-linearity

RESEARCH REPORT No. RA-17

UDC 621.391.812.63:621.391.827.4 1968/19

THE BRITISH BROADCASTING CORPORATION
ENGINEERING DIVISION

RESEARCH DEPARTMENT

IONOSPHERIC NON-LINEARITY

Research Report No. RA-17
UDC 621.391.812.63: 1968/19
621.391.827.4

P. Knight, B.A., M.I.E.E.

Head of Research and Development

This Report is the property of the British Broadcasting Corporation and may not be reproduced in any form without the written permission of the Corporation.

This Report uses SI units in accordance with B.S. document PD 5686.

Research Report No. RA-17

IONOSPHERIC NON-LINEARITY

Section Title Page

SUMMARY 1

LIST OF PRINCIPAL SYMBOLS 1

1. INTRODUCTION 1

2. THE MECHANISM OF IONOSPHERIC NON-LINEARITY 2

3. INCREASE IN COLLISION FREQUENCY DUE TO A DISTURBING WAVE 3

3.1. Increase in Collision Frequency when the Earth's Magnetic Field is Neglected 3
3.2. Increase in Collision Frequency, taking into account the Effect of the Earth's Magnetic Field 4

4. CHANGE IN ABSORPTION OF WANTED WAVE DUE TO INCREASE OF COLLISION FREQUENCY 5

5. THE NIGHT-TIME BEHAVIOUR OF THE IONOSPHERE AT MEDIUM FREQUENCIES . . . 5

5.1. Non-linearity in the D-Layer 7
5.2. Non-linearity in the E-Layer 7
5.3. Non-linearity on a Typical Path 7

6. AUDIO-FREQUENCY DISTORTION CAUSED BY IONOSPHERIC NON-LINEARITY 9

6.1. Cross-modulation 10
6.2. Self-distortion 11
6.3. Interaction between Two Strong Waves 11

7. CONCLUSIONS 12

8. REFERENCES 13

May 1968

Research Report No. RA-17
UDC 621.391.812.63: 1968/19
621.391.827.4

IONOSPHERIC NON-LINEARITY

SUMMARY

The mechanism which causes the ionosphere to behave as a non-linear transmission medium when illuminated by a strong radio wave is discussed. It is shown that the non-linear effect is accentuated near the gyromagnetic frequency if the extraordinary wave is strongly-excited by the transmitter. Calculation of the magnitude of the non-linear effect by means of a ray-tracing technique is described and expressions are derived which enable the attenuation, distortion and cross-modulation caused by non-linearity to be calculated.

LIST OF PRINCIPAL SYMBOLS

A	Total ionospheric attenuation of wave (dB)
A_U	Total ionospheric attenuation of unmodulated wave (dB)
A_0	Total ionospheric attenuation of wave in absence of disturbing wave (dB)
a	Attenuation per unit path length of wave (dB/km)
a_0	Attenuation per unit path length of wave in absence of disturbing wave
$\hat{E}$	Peak field strength of wave (mV/m)
E_p	Plasma field strength, defined in Equation (5) (mV/m)
$\hat{E}_U$	Peak field strength of unmodulated wave (mV/m)
e	Charge of electron
G	A constant defined in Equation (2)
M	Modulation index of wanted wave
M'	Modulation index of unwanted wave
m	Mass of electron
N	Number of electrons per unit volume
V	Carrier amplitude of received signal
V_0	Carrier amplitude of received signal in absence of disturbing wave
v	Average of random electron velocities
v_0	Average of random electron velocities in absence of disturbing wave
z	Distance measured along path of wave (km)
$\delta A(t)$	Instantaneous increase in attenuation due to ionospheric non-linearity
δA_U	Increase in attenuation of unmodulated wanted wave due to ionospheric non-linearity
$\delta A'_U$	Increase in attenuation of unmodulated unwanted wave due to ionospheric non-linearity
δA_F	Reduction of the fundamental component of the wanted wave modulation (dB)
$\delta A'_F$	Reduction of the fundamental component of the unwanted wave modulation (dB)
ν	Electron collision frequency
ν_0	Electron collision frequency in the absence of disturbing wave
ϕ_1	$\tan^{-1}(\omega_m/G\nu_0)$
ϕ_2	$\tan^{-1}(2\omega_m/G\nu_0)$
ψ	Angle between direction of wave and Earth's magnetic field
ω	Angular frequency of wave
ω_H	Angular gyromagnetic frequency
ω_m	Angular modulation frequency of wanted wave
ω'_m	Angular modulation frequency of unwanted wave

1. INTRODUCTION

Effects caused by non-linearity in the ionosphere are becoming increasingly important with the trend towards the use of higher powers for night-time sky-wave broadcasting on low and medium frequencies. These effects may be classified as follows:

(i) the modulation of a powerful disturbing wave may be impressed on another wave traversing the same region of the ionosphere, so causing interference. This is the so-called Luxembourg effect, first observed in 1933.

(ii) the ionospheric absorption of a strong wanted wave may be increased and harmonic distortion of the audio-frequency modulation may occur.

(iii) when two strong waves are radiated simultaneously, intermodulation may take place in addition to cross-modulation.

There is considerable evidence, both theoretical and experimental, which suggests that these effects may be enhanced if the frequency of the disturbing wave is close to the gyromagnetic frequency.

The general theory of ionospheric non-linearity is surveyed in this report and extended to the situation which arises in tropical latitudes, where the extraordinary wave is excited strongly. Use is made of a simple ray-tracing technique to assess the effect of non-linearity on an incident wave which excites the ordinary and extraordinary waves in roughly equal proportions. The dependence of distortion and cross-modulation on the modulation frequency is discussed and the theory extended to include effects occurring when two strong waves illuminate the same region of the ionosphere.

2. THE MECHANISM OF IONOSPHERIC NON-LINEARITY

The ionosphere consists of a mixture of neutral gas molecules, positive ions and free electrons. The electrons, being relatively light, are easily deflected by electric and magnetic fields. When a wave propagates through the ionosphere its electric field imparts an oscillatory motion to the electrons, which re-radiate like miniature aerials. The re-radiation modifies the velocity of propagation of the wave and refraction occurs when the electron density varies, so that if the frequency is not too high waves are refracted back towards the Earth.

If the Earth's magnetic field were absent the oscillatory motion of the electrons would be parallel to the direction of the electric field of the incident wave and the re-radiation would have the same polarization as the incident wave; the wave would therefore traverse the ionosphere without change of polarization. The Earth's magnetic field modifies the oscillatory motion of the electrons, causing them to move in complicated orbits. Their re-radiation is not of the same polarization as the incident wave and consequently the polarization of the resultant wave changes continually as it traverses the ionosphere.

There are, however, two incident-wave polarizations (in general elliptical) which excite the electrons in such a way that the waves re-radiated by the electron have the same polarization; these waves therefore traverse the ionosphere without change of polarization. These two special polarizations are termed the 'ordinary' and 'extraordinary' waves, the parameters describing their polarizations depending on the angle between the direction of propagation and the direction of the Earth's magnetic field. Since these two waves traverse the ionosphere with different velocities and attenuations it is usual to resolve the incident wave into the two components and then consider them separately.

Both the ordinary and extraordinary waves impart a minute but significant quantity of energy to each electron. Now the constituents of the ionosphere are in continuous random motion and collisions between electrons and heavier particles occur frequently. When such a collision occurs the ordered energy supplied by the wave is converted into random energy and more energy has to be supplied by the wave to set the electron in motion again; the wave therefore suffers loss and is attenuated. The more frequently the electrons collide the greater is the attenuation, unless collisions occur so often that the electrons have insufficient time to acquire the full amount of ordered energy which the wave can supply before the next collision takes place.

Up to this point we have been considering waves which are so weak that the oscillatory motion which they impart to the electrons is small compared with the random motion. If the wave is sufficiently strong, however, the random motion of the electron may be increased significantly by transfer of energy from the oscillatory motion at each collision. The chances of a collision occurring are then greater, so that the collision frequency is increased and this may result in an increase of absorption, as discussed in the previous paragraph. This increased loss would apply not only to the strong wave but to all other waves traversing the same region of the ionosphere, however weak they might be. If the strong wave (which will be referred to as the disturbing wave) is amplitude-modulated at a sufficiently low audio frequency the attenuation will vary at audio frequency and the modulation of the disturbing wave will be superimposed on the weaker wave; this is the basis of the Luxembourg effect. If the disturbing wave itself is being received, a non-linear relationship will exist between the received and transmitted powers and its waveform will be distorted as a result.

Special consideration must be given to the situation which arises when the disturbing wave excites the extraordinary wave. Near the gyromagnetic frequency a relatively weak extraordinary wave can impart a large ordered motion to the electrons and cause a significant increase in collision frequency. This gyro-resonance effect was proposed theoretically by Bailey[1,2] in 1937 and has since been demonstrated experimentally[3,4]. Here we consider the situation which arises if the disturbing wave is the extraordinary component of a wave whose frequency is near the gyromagnetic frequency, the wanted wave being the ordinary component. Ordinary and extraordinary waves are excited in roughly equal proportions on East-West paths at magnetic dip latitudes of about 45° and on long North-South paths near the magnetic equator.

To determine the influence of a disturbing wave on a wanted wave, the increase in collision frequency due to the disturbing wave is first calculated. The change in the absorption coefficient of the wanted wave is then found; this yields a factor by which the attenuation of the wanted wave must must be multiplied. Provided the attenuation of the wanted wave in the absence of the disturbing wave is known, the change in attenuation can be found. It is then possible to calculate the effect which the attenuation change has on the modulation of the wanted wave; the cross-modulation impressed on other waves may also be determined.

The theory of ionospheric non-linearity is discussed in more detail in the sections which follow.

3. INCREASE IN COLLISION FREQUENCY DUE TO A DISTURBING WAVE

3.1. Increase in Collision Frequency when the Earth's Magnetic Field is Neglected

Since the constituents of the ionosphere are in continuous random motion, collisions frequently occur. If we follow the movement of a single electron we find that, on average, ν collisions occur every second even though the intervals between successive collisions are irregular.* In discussing the behaviour of the ionosphere the effective collision frequency is assumed to be slightly greater than the actual collision frequency, for reasons which are beyond the scope of this report. In all that follows, ν will be assumed to denote the effective collision frequency.

If the Earth's magnetic field is neglected it may be shown that the work done between collisions by the wave in setting an electron in motion is, on average, given by

$$w = \frac{1}{2}\frac{e^2}{m}\frac{\hat{E}^2}{\nu^2 + \omega^2} \qquad (1)$$

per collision, where

e and m are the charge and mass of the electron, respectively

$\hat{E}$ is the peak value of the incident electric field

ν is the collision frequency

ω is the angular wave frequency

When a collision occurs this energy is converted into random energy and is shared between the electron and the heavy particle with which it collides. The mean random velocity of the electrons therefore increases from v_0 in the absence of the wave to v in the presence of the wave. When a steady state has been reached the average energy of an electron is increased by an amount given by

$$\tfrac{1}{2}\, m(v^2 - v_0^2)G = w \qquad (2)$$

where G is a constant which has been determined experimentally** and is approximately equal to $1{\cdot}7 \times 10^{-3}$.

Now the collision frequency ν is directly proportional to v because the mean free path traversed by the electrons between collisions is independent of their velocity. From Equations (1) and (2) we may therefore obtain the following result

$$\left(\frac{\nu}{\nu_0}\right)^2 = \left(\frac{v}{v_0}\right)^2 = 1 + \left(\frac{e}{m}\right)^2 \frac{1}{v_0^2 G} \cdot \frac{\hat{E}^2}{\nu^2 + \omega^2} \qquad (3)$$

where ν_0 is the collision frequency in the absence of the incident wave. Equation (3) shows that ν increases in the presence of the wave.

Equation (3) may be written in the form

$$\nu^2 = \nu_0^2\left[1 + \frac{\nu_0^2 + \omega^2}{\nu^2 + \omega^2}\left(\frac{\hat{E}}{E_p}\right)^2\right] \qquad (4)$$

where

$$E_p = v_0\,\frac{m}{e}\left[G(\nu_0^2 + \omega^2)\right]^{1/2} \qquad (5)$$

The quantity E_p has the dimensions of field strength and is known as the 'plasma field strength'; in the D- and E-layers it is given approximately by

$$E_p \simeq 150\left[\left(\frac{\nu_0}{2\pi}\right)^2 + f^2\right]^{1/2} \text{ mV/m} \qquad (6)$$

where $\nu_0/2\pi$ and f are in MHz. At heights greater than 80 km, $\omega^2 \gg \nu_0^2$ throughout the m.f. band; the expression for E_p then simplifies to

$$E_p \simeq 150f \text{ mV/m}$$

In the m.f. band, $\hat{E}/E_p$ is unlikely to exceed unity even with the highest powers contemplated.

---

* In the D- and E-regions, where most absorption takes place, collisions with neutral molecules occur much more frequently than collisions with positive ions. If an electron collides with a positive ion its free existence may be terminated and it then plays no further part in the propagation of the wave. The concept of an average collision frequency is, however, still valid for the free electrons which remain.

**There is some uncertainty about the correct value for G and its value cannot be derived from simple theoretical considerations. For a discussion of this problem see Appendix II of Reference 4.

Consequently the increase in ν is small compared with ω and Equation (4) may be simplified to

$$\nu^2 = \nu_0^2\left[1 + (\hat{E}/E_p)^2\right]$$

with little error; this approximation was made by Huxley and Ratcliffe.[5] In Section 3.2 we shall show that a modified form of Equation (4) applies when the Earth's magnetic field is taken into consideration and that the quantity equivalent to $\hat{E}/E_p$ may be appreciably greater than 1. The approximate solution of Huxley and Ratcliffe is no longer valid and we must use the exact solution of Equation (4), which is a quadratic in ν^2. For convenience the exact solution will be stated here; it is

$$\nu^2 = \frac{1}{2}\left\{\nu_0^2 - \omega^2 + (\nu_0^2 + \omega^2)\left[1 + \frac{4\nu_0^2}{\nu_0^2 + \omega^2}\left(\frac{\hat{E}}{E_p}\right)^2\right]^{1/2}\right\} \tag{7}$$

This result has been given in slightly different form by Ginzburg and Gurevich.[6] Curves showing the relationship between ν/ν_0 and $\hat{E}/E_p$ derived from Equation (7) are shown in Fig. 1.

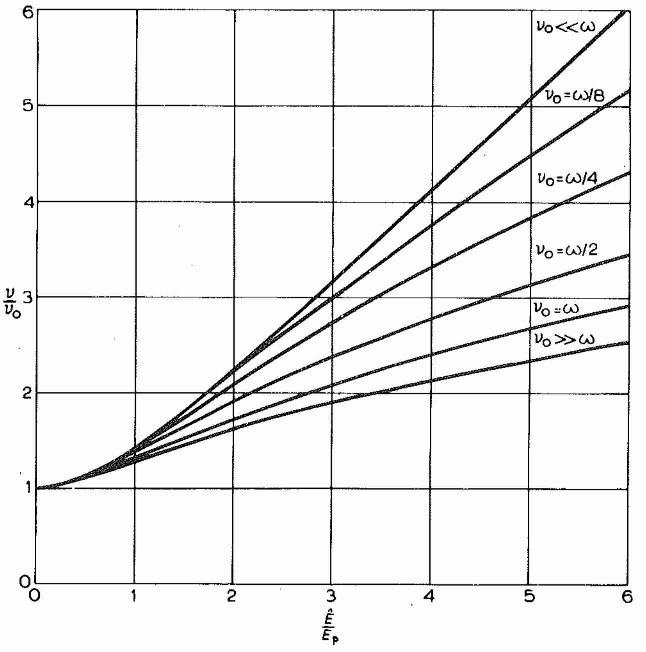


Fig. 1 - Increase of collision frequency as a function of strength of disturbing wave
Earth's magnetic field neglected

It should be noted that the values of ν/ν_0 given by Equations (3) and (7) are 'steady-state' values, applicable when E is constant, i.e. when the disturbing wave is unmodulated. If the wave is amplitude modulated $\hat{E}$ varies with time and may be written

$$\hat{E} = \hat{E}_U\,(1 + M\cos\omega_m t)$$

where $\hat{E}_U$ is the peak value of the unmodulated wave, M the modulation index and ω_m the angular modulation frequency. When Huxley and Ratcliffe's approximation is valid the ionosphere may be represented by the electrical analogue model illustrated in Fig. 2. This shows that ν^2 rises and falls with time constant $1/G\nu_0$, in the E-layer approximately equal to 1·25 milliseconds. At very low modulation frequencies, ν varies over the modulation cycle in accordance with the 'steady-state' solution. At high modulation frequencies, however, only the constant term in the expression for $\hat{E}^2$ is important; this is equal to $\hat{E}_U^2(1 + M^2/2)$, which is proportional to the total power contained in the carrier and sidebands. Thus at high audio frequencies ν remains constant over the modulation cycle, its value depending only on the power density of the incident wave.

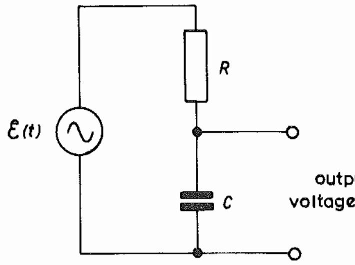


Fig. 2 - Analogue model of the ionosphere

Time constant $CR = 1/G\nu_0$

$$\text{Input voltage, } \mathcal{E}(t) = \nu_0^2\left[1 + \left(\frac{\hat{E}}{E_p}\right)^2\right]$$

$$= \nu_0^2\left[1 + \left(\frac{\hat{E}_U}{E_p}\right)^2\left(1 + \frac{M^2}{2} + 2M\cos\omega_m t + \frac{M^2}{2}\cos 2\omega_m t\right)\right]$$

The effect which the variation of ν at audio frequency has on the received signal is discussed fully in Section 6.

3.2. Increase in Collision Frequency, taking into account the Effect of the Earth's Magnetic Field

When the effect of the Earth's magnetic field is taken into consideration the average energy acquired by the electrons between collisions is modified and Equation (1) applies only when the electric vector of the incident wave is parallel with the direction of the Earth's magnetic field. The theory contained in Section 3.1 must therefore be restated with a modified value of w.

It has been shown by Bailey[2,7] that the average work done by a plane-polarized incident wave on an electron is

$$w = \frac{e^2\hat{E}^2}{4m}\left\{\left[\frac{1}{\nu^2 + (\omega - \omega_H)^2} + \frac{1}{\nu^2 + (\omega + \omega_H)^2}\right]\sin^2\psi + \frac{2\cos^2\psi}{\nu^2 + \omega^2}\right\} \tag{8}$$

where ω_H is the angular gyromagnetic frequency and ψ is the angle between the directions of the electric field of the incident wave and the Earth's magnetic field. We may compare this equation with Equation (1); as expected, Equation (8) reduces to Equation (1) when $\psi = 0$, since the Earth's magnetic field then has no effect on the motion of the electrons. When $\psi \neq 0$, the first term of Equation (8) is large at frequencies near the gyromagnetic frequency, because of the resonant motion of the electrons.

We now proceed as in Section 3.1, replacing Equation (1) by Equation (8). The complete expression for ν/ν_0 leads to a quartic in ν^2. Bailey has obtained a solution valid for relatively weak disturbing waves by replacing ν in Equation (8) by ν_0; this is equivalent to the approximation of Huxley and Ratcliffe. There is, however, less justification for making the approximation in this case because large increases in ν can occur near the gyromagnetic frequency.

A new approach to the problem is adopted here by restricting the calculation of ν/ν_0 to $\psi = 0$ and $\psi = 90°$ only.

When $\psi = 0$ the result given in Section 3.1 applies. When $\psi = 90°$ the last term on the right-hand side of Equation (8) is zero and it can be shown that, in the m.f. band, the second term in the square bracket is always small compared with the first; the second term will therefore be neglected. Substitution of the resulting value of w in Equation (2) then leads to the following quadratic in ν^2, equivalent to Equation (3)

$$\left(\frac{\nu}{\nu_0}\right)^2 = 1 + \frac{1}{2}\left(\frac{e}{m}\right)^2 \frac{1}{v_0^2 G} \cdot \frac{\hat{E}^2}{\nu^2 + (\Delta\omega)^2} \qquad (9)$$

where $\Delta\omega = |\omega - \omega_H|$, the deviation from the gyromagnetic frequency.

We now define a 'modified plasma field-strength' E'_p as

$$E'_p = v_0 \frac{m}{e}\left\{2G\left[\nu_0^2 + (\Delta\omega)^2\right]\right\}^{1/2} = E_p\left\{\frac{2[\nu_0^2 + (\Delta\omega)^2]}{\nu_0^2 + \omega^2}\right\}^{1/2} \qquad (10)$$

Equation (9) may then be written as follows

$$\nu^2 = \nu_0^2\left[1 + \frac{\nu_0^2 + (\Delta\omega)^2}{\nu^2 + (\Delta\omega)^2}\left(\frac{\hat{E}}{E'_p}\right)^2\right] \qquad (11)$$

Comparison of Equations (4) and (11) shows that they are of the same form. Consequently ν/ν_0 may be determined from Fig. 1 if E_p is replaced by E'_p and ω by $\Delta\omega$. The equivalent circuit of Fig. 2 may still be used to estimate how ν varies with modulation frequency provided E_p is replaced by E'_p.

In the D-layer there is little difference between the values of ν/ν_0 for $\psi = 0$ and 90° and therefore no great uncertainty about the value of ν/ν_0 for other values of ψ. In the E-layer it is convenient to calculate ν/ν_0 for ordinary and extraordinary disturbing waves separately; for transverse propagation the two waves correspond to $\psi = 0$ and $\psi = 90°$ respectively. The value of ν/ν_0 due to the combined effect of the two disturbing waves cannot easily be calculated; the larger of the values for the individual waves is used here as a basis for further calculation.

4. CHANGE IN ABSORPTION OF WANTED WAVE DUE TO INCREASE OF COLLISION FREQUENCY

The relationship between absorption and collision frequency is complicated and has been discussed fully in Chapters 9 and 10 of Reference 8. At medium frequencies, however, certain simplifying assumptions may be made for paths involving oblique incidence, since most absorption occurs at levels where the refractive index does not differ greatly from unity. Here the attenuation of the ordinary wave is proportional to $\nu/(\nu^2 + \omega^2)$ provided the direction of propagation is perpendicular to the direction of the Earth's magnetic field. The rate of attenuation a of a wanted wave is then related to its rate of attenuation a_0 in the absence of the disturbing wave by the expression

$$\frac{a}{a_0} = \frac{\nu}{\nu_0} \cdot \frac{\nu_0^2 + \omega^2}{\nu^2 + \omega^2} = \frac{\nu}{\nu_0} \cdot \frac{1 + (\omega/\nu_0)^2}{(\nu/\nu_0)^2 + (\omega/\nu_0)^2} \qquad (12)$$

Thus if ν/ν_0 is known, a/a_0 can be determined.

Fig. 3 shows how a/a_0 varies as ν/ν_0 increases. It will be seen that if ν_0 is greater than ω the attenuation always decreases, for reasons which were discussed in Section 2. If ν_0 is less than ω the attenuation increases until $\nu = \omega$, after which it falls.

5. THE NIGHT-TIME BEHAVIOUR OF THE IONOSPHERE AT MEDIUM FREQUENCIES

Although the theory discussed in the previous sections enables the fractional increase in attenuation caused by disturbing wave to be calculated, it is of no practical use unless the attenuation which occurs in the absence of the disturbing wave is known. To assess this, the properties of the ionosphere must be studied more closely.

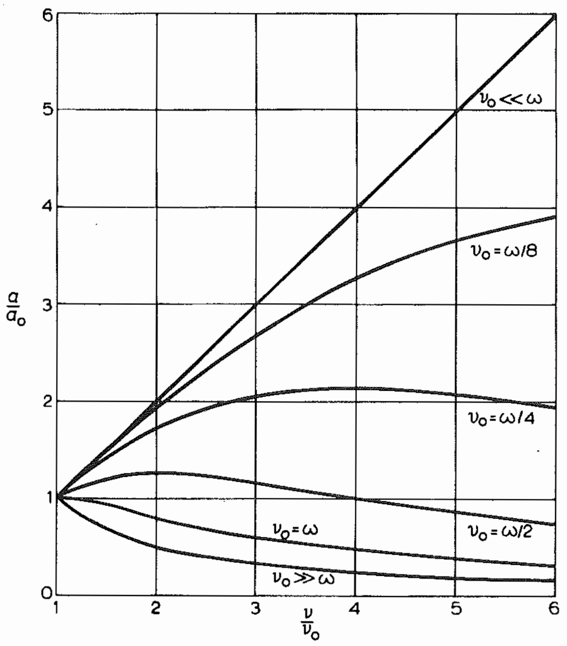


Fig. 3 - Change of attenuation of wanted wave as a function of increase in collision frequency

Transverse propagation of the ordinary wave

Figs. 4(*a*) and (*b*) show typical night-time electron density and collision frequency profiles for the D- and E-layers, taken from Figs. 1 and 7, curve (*b*), of Reference 9 but with the maximum electron density increased to 2000 per cm^3 as this value is believed to be more accurate.[10] Fig. 4(*c*) shows the refractive index for transverse propagation of the ordinary wave at 1 MHz. It will be seen that the refractive index is very close to unity in the D-layer and does not fall below 0·915 in the E-layer. Thus 1 MHz waves incident at angles less than arc sin 0·915 (66°) will penetrate the E-layer at night; the skip distance is 440 km.

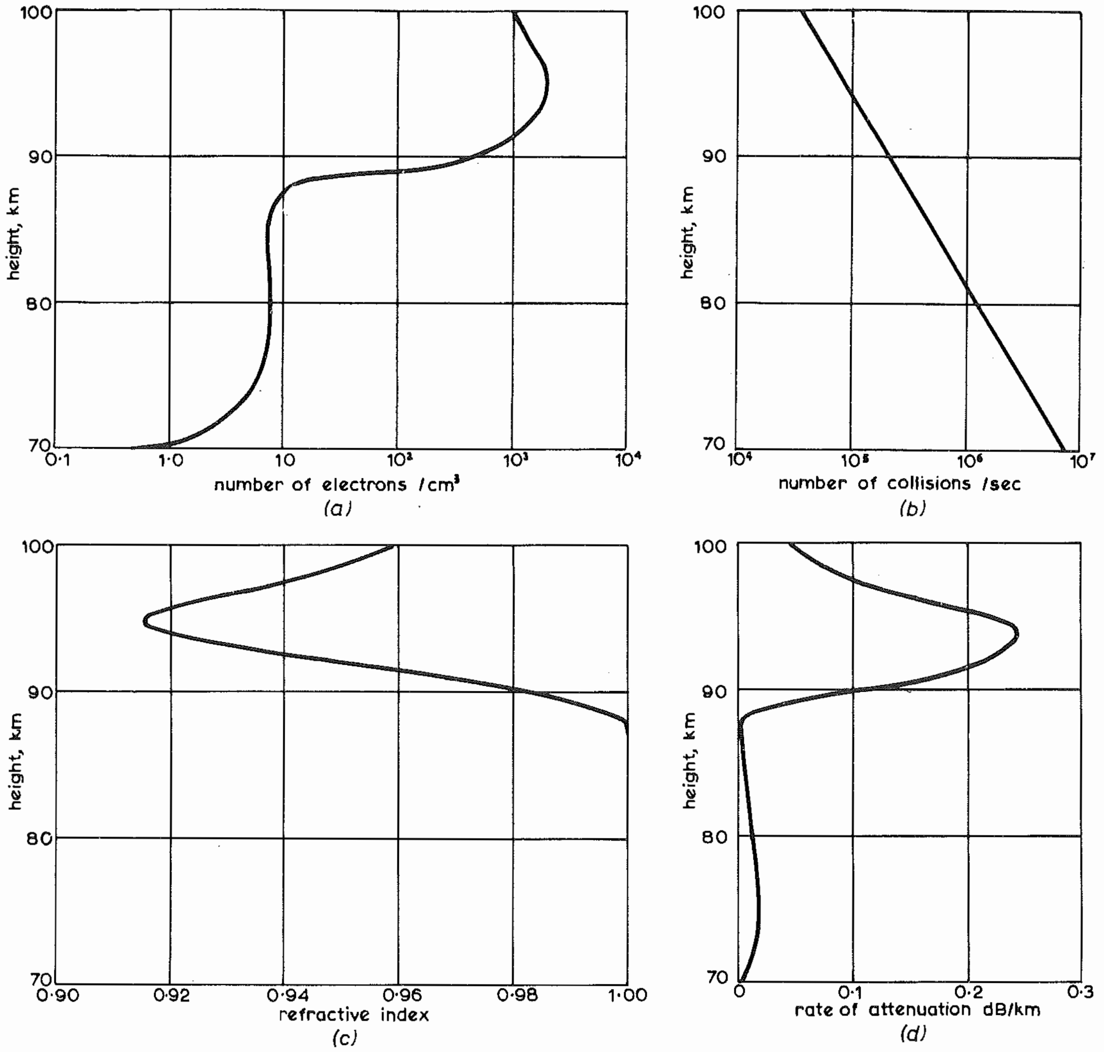


Fig. 4 - Night-time ionospheric parameters

(*a*) Electron density (*b*) Effective collision frequency

(*c*) Refractive index for transverse propagation of the ordinary wave at 1 MHz

(*d*) Vertical-incidence attenuation for transverse propagation of the ordinary wave at 1 MHz (one traversal of the layer)

The rate of attenuation of a vertically-incident 1 MHz ordinary wave is shown in Fig. 4(*d*). It is approximately proportional to $N\nu$ where N is the electron density, and rises to a maximum, as N increases and ν decreases, at about 74 km. This occurs in the D-layer, which at night should be regarded more as a region of enhanced absorption than as a region of increased electron density. The sudden rise in electron density in the E-layer causes the rate of attenuation to rise to a second and greater maximum at about 94 km, despite the greatly reduced collision frequency at this height.

In calculating the effect of a disturbing wave it is important to distinguish between D- and E-layer attenuation, because ν_0 differs by a factor of more than 20 between the two layers. The calculation of D-layer loss presents no difficulty because at m.f. the wave is not deviated significantly and the path may be assumed to be a straight line through the layer, even at oblique incidence. The E-layer loss is not so easy to determine because it depends on the path length and depth of penetration in the layer. A crude estimate may, however, be made from propagation curves based on measured field strengths and published by the EBU[11] by subtracting the calculated D-layer loss from the total ionospheric attenuation. For more accurate estimates use must be made of ray-tracing techniques, as described in Section 7.

Fig. 5 shows D-layer losses calculated from the profiles of Fig. 4 and E-layer losses estimated from the EBU curves, and shows how these losses vary with frequency and distance. They represent median values for night time. Although the D-layer attenuation decreases as the frequency rises the E-layer losses show the opposite trend; despite the decreased rate of attenuation at the higher frequencies, waves penetrate more deeply into the E-layer and consequently suffer a greater total attenuation.

5.1. Non-linearity in the D-Layer

The absorption of the wanted wave is greatest at 74 km, where the collision frequency ν_0 is 3·7 x 10^6 per sec. Thus $\nu_0 = \omega$ at about 0·6 MHz. Fig. 3 shows that an increase in ν will cause the absorption to decrease at frequencies below 0·6 MHz and suggests that a/a_0 is unlikely to exceed 1·3 at any frequency in the m.f. band. The highest values of a/a_0 occur at the highest frequencies, but here the total D-layer loss is least, as Fig. 5 shows. We therefore conclude that non-linearity effects are unimportant in the D-layer; this conclusion has been confirmed by detailed calculation for an 800 km path.

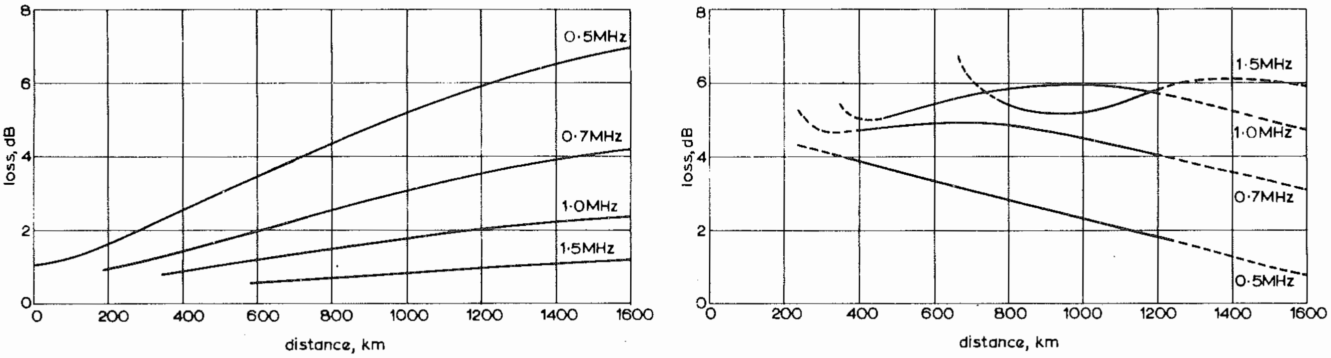
5.2. Non-linearity in the E-Layer

On entering the D-layer the incident wave may be resolved into ordinary and extraordinary waves which suffer different attenuations. When the ordinary wave reaches the E-layer it may still be sufficiently strong to act as a disturbing wave. The extraordinary wave may also act as a disturbing wave even though it is more heavily attenuated in the D-layer because of gyro-frequency resonance. Here the effect of each wave is considered separately because the theory for assessing their combined effects has not yet been developed.

In the E-layer at m.f. $\omega \gg \nu_0$ and consequently a/a_0 may be assumed to be equal to ν/ν_0; this result follows from Equation (12).

5.3. Non-linearity on a Typical Path

Although non-linearity is important only in the E-layer, allowance must be made for attenuation in



(*a*) (*b*)

Fig. 5 - Variation of ionospheric attenuation with distance and frequency

(*a*) D-layer loss (*b*) E-layer loss

the D-layer. Fig. 6 shows vertical-incidence losses for ordinary and extraordinary waves, calculated by integrating curves similar to Fig. 4(*d*). If these losses are divided by the cosine of the angle of incidence, the strengths of both waves at the base of the E-layer may be calculated. Values of ν/ν_0 and hence a/a_0, may then be calculated for the base of the E-layer, using the method described in Section 3.

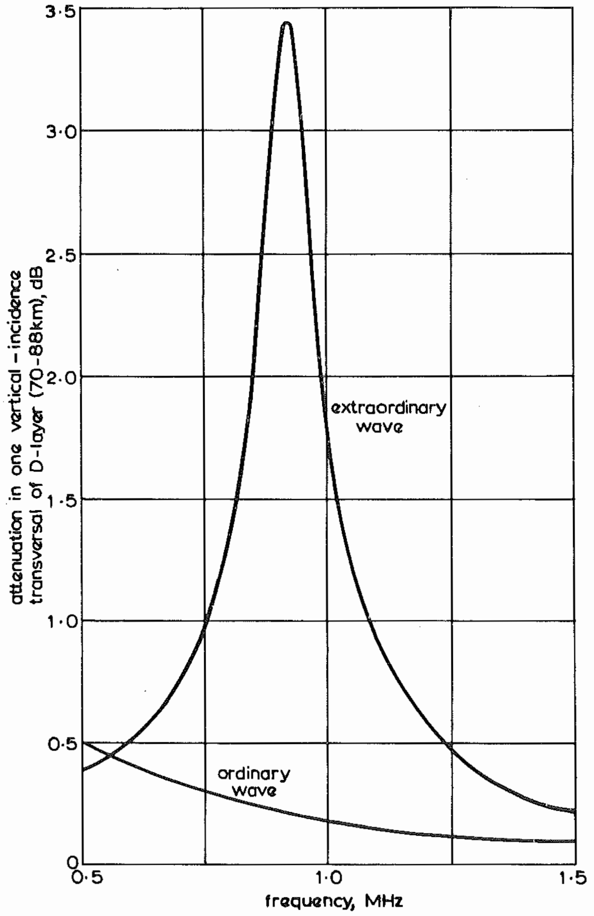


Fig. 6 - Vertical-incidence attenuation in the D-layer
Transverse propagation. Gyromagnetic frequency 920 kHz.

Fig. 7 shows the values of a/a_0 which would occur at the base of the E-layer at distances of 400 and 800 km from transmitters having effective radiated powers (e.r.p.s) of 7½ and 30 MW. The unbroken curves are the values of a/a_0 occurring when the transmitters excite only the transverse ordinary wave; these curves apply to propagation in all directions near the magnetic poles and are independent of the gyromagnetic frequency. The unbroken curves are approximately correct for European latitudes since the excitation of the extraordinary wave by a vertical aerial is relatively weak.

The broken curves show the values of a/a_0 occurring if the transmitters excite the transverse extraordinary wave strongly; this situation would arise when transmitting on an East-West path near the magnetic equator. An interesting feature of the broken curves is the double-hump shape centred on the gyromagnetic frequency, the central minimum being due to the high absorption loss suffered by the wave in passing through the D-layer. It is of interest to note that double-hump curves with roughly the same frequency separation between maxima have been observed in cross-modulation experiments by Cutolo in Europe[3] and Hibberd in Australia.[4]

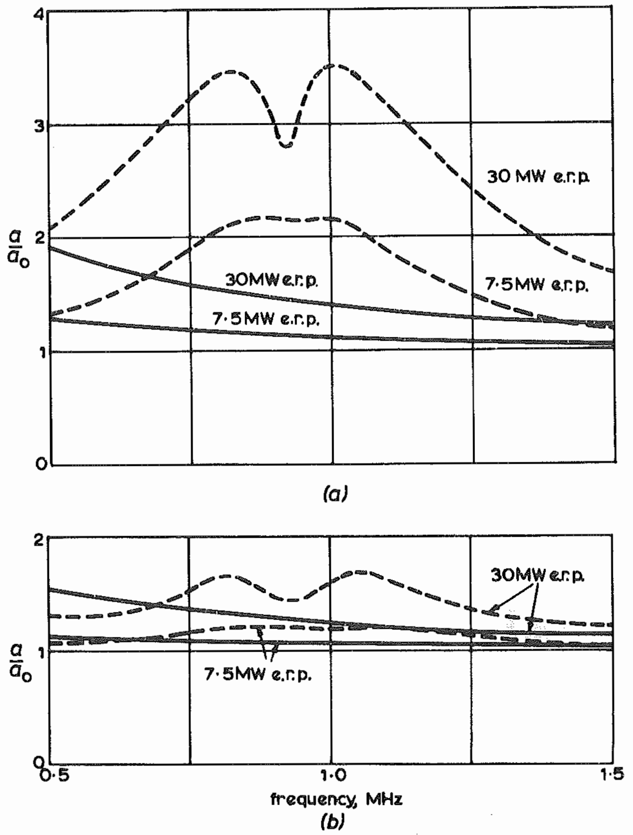


Fig. 7 - Increase in attenuation of ordinary wave at base of E-layer
(*a*) Reflection point 400 km from transmitter
(*b*) Reflection point 800 km from transmitter

——— Transverse ordinary disturbing wave
– – – Transverse extraordinary disturbing wave; gyromagnetic frequency 920 kHz

The large values of a/a_0 due to the extraordinary wave apply only at the point where the wanted wave enters and leaves the E-layer, because the extraordinary wave is rapidly attenuated above the base of the layer. Furthermore, for frequencies above the gyromagnetic frequency, the extraordinary wave is reflected at a lower level than the ordinary wave and never reaches the upper part of the ordinary-wave path, where most of the ordinary-wave attenuation occurs; in this region, therefore, any non-linear effects are caused by the ordinary wave alone.

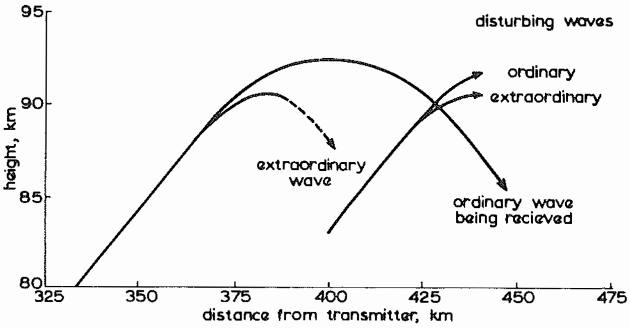


Fig. 8 - Wave paths in the ionosphere

Frequency 1·3 MHz. Total path length 800 km

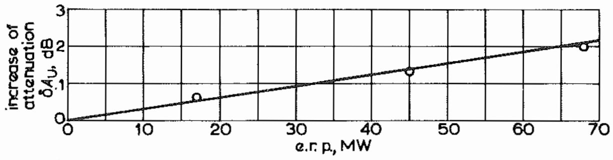


Fig. 9 - Increase of attenuation due to non-linearity

o Calculated values

40° dip latitude; 800 km path; Frequency 1·1 to 1·5 MHz; Gyromagnetic frequency 920 kHz

The overall increase in the attenuation of the ordinary wave radiated from a high-power transmitter must therefore be calculated by tracing the path of the wave through the layer and finding the attenuation increase at each point. This may be done by regarding the E-layer as a set of uniform horizontal slabs each 1 km thick and using Snell's Law to trace the ray paths. As an example, Fig. 8 shows paths calculated in this way, drawn as smooth curves with an exaggerated vertical scale. The Appleton-Hartree formula[9] was used to calculate the refractive index and absorption index in each layer, transverse propagation being assumed throughout.

The ordinary and extraordinary waves whose paths are shown in Fig. 8 were assumed to be excited in roughly equal proportions by the transmitter; this situation would arise, for example, near 45° dip latitude on East-West paths. The attenuation of the ordinary wave, when a low-power transmitter is used, was calculated by adding attenuations calculated from the Appleton-Hartree formula for individual layers. The additional attenuation resulting from the use of a high-power transmitter was then estimated as follows. The strengths of the ordinary and extraordinary interfering waves in each layer were calculated and the values of ν/ν_0 and hence a/a_0, which each would cause individually were determined from Fig. 1; the larger of the two values was adopted where both waves are present because of the difficulty, mentioned earlier, of calculating the combined effect of the two waves. The weak-signal attenuations previously calculated for the ordinary wave for each layer were multiplied by the appropriate values of a/a_0 and again added, so giving the total increase in attenuation due to non-linearity. It is worth noting that the attenuation increase suffered by the wave on its downward path is similar to the increase on the upward path even though the wave has been attenuated in the E-layer, because other upgoing waves, shown in Fig. 8, disturb the downcoming wave.

Calculations were made for an 800 km path at frequencies of 1·1, 1·3 and 1·5 MHz. The calculated weak-signal E-layer attenuations at these frequencies are 6·9, 7·7 and 13·1 dB respectively; these losses are consistent with those estimated from the EBU curves, shown in Fig. 5(*b*), the high value at 1·5 MHz arising because the ordinary wave almost penetrates the layer at this frequency. The attenuation increases due to non-linearity were calculated for three values of e.r.p. and were found to be similar on all three frequencies. Fig. 9 shows average values for the three frequencies plotted as a function of e.r.p.; it will be seen that the relationship between the e.r.p. and the attenuation increase is approximately linear for the range of powers considered. On theoretical grounds a linear relationship would not be expected unless Huxley and Ratcliffe's approximation were valid. The fact that the calculation described in this section leads to a linear relationship shows that Huxley and Ratcliffe's approximation is justified for the path as a whole even though it may not be valid where the wave enters and leaves the E-layer. The calculations showed that the greater part of the ordinary-wave attenuation occurs near the apex of the trajectory and here the approximation is valid for the powers considered.

6. AUDIO-FREQUENCY DISTORTION CAUSED BY IONOSPHERIC NON-LINEARITY

In this section we consider how the modulation of the wanted wave is affected by the disturbing wave. Huxley and Ratcliffe's approximation will be used, since it has been shown in Section 5 to be valid for a typical path taken as a whole.

Suppose that z represents distance measured along the path of the wanted wave and that $a_0(z)$ is its rate of attenuation, in dB per unit path length, in the absence of the disturbing wave. The total attenuation is therefore $A_0 = \int a_0(z)\mathrm{d}z$.

When the disturbing wave is present the rate of attenuation increases to $a(z)$. We may therefore write

$$\frac{a(z)}{a_0(z)} \simeq \frac{\nu(z)}{\nu_0(z)} \simeq \left[1 + \left(\frac{\hat{E}(z)}{E_\mathrm{p}(z)}\right)^2\right]^{½} \simeq 1 + \frac{1}{2}\left(\frac{\hat{E}(z)}{E_\mathrm{p}(z)}\right)^2 \tag{13}$$

provided $\hat{E}(z)$ varies very slowly with time. In Equation (13), $\hat{E}(z)$ is the peak value of the disturbing wave and $E_\mathrm{p}(z)$ the plasma field strength, or the modified plasma field strength in the case of an extraordinary disturbing wave. Both quantities vary with height and are therefore functions of z.

If $\hat{E}(z)$ varies more rapidly with time, the instantaneous rate of attenuation depends also on the ionospheric time constant $1/G\nu_0$. If

$$\hat{E}(z) = \hat{E}_U(z)\,[1 + M\cos\omega_m t]$$

then the electrical analogue model of Fig. 2 shows that ν^2 is given by

$$\nu^2 = \nu_0^2\left[1+\left(\frac{\hat{E}_U}{E_p}\right)^2 f(t)\right]$$

where
$$f(t) = 1 + \frac{M^2}{2} + 2M\cos\phi_1\cos(\omega_m t-\phi_1) + \frac{M^2}{2}\cos\phi_2\cos(2\omega_m t-\phi_2) \qquad (14)$$

$$\phi_1 = \tan^{-1}(\omega_m CR) = \tan^{-1}(\omega_m/G\nu_0)$$

$$\phi_2 = \tan^{-1}(2\omega_m CR) = \tan^{-1}(2\omega_m/G\nu_0)$$

In the E-layer the attenuation increase is therefore given by

$$\frac{a(z)}{a_0(z)} = \frac{\nu(z)}{\nu_0(z)} = \left[1+\left(\frac{\hat{E}_U(z)}{E_p(z)}\right)^2 f(t)\right]$$

$$\simeq 1 + \frac{1}{2}\left(\frac{\hat{E}_U(z)}{E_p(z)}\right)^2 f(t) \qquad (15)$$

Integration of Equation (15) gives the total instantaneous attenuation of the wanted wave, as follows:

$$A(t) = \int a(z)\,dz = \int a_0(z)\,dz + \tfrac{1}{2}f(t)\int a_0(z)\left(\frac{\hat{E}_U(z)}{E_p(z)}\right)^2 dz \qquad (16)$$

The first term on the right hand side of Equation (16) is the attenuation which would occur in the absence of the disturbing wave and the second term is the increase due to non-linearity.* Since $f(t) = 1$ when the disturbing wave is unmodulated, the second term may be written in the form

$$\delta A(t) = f(t)\,\delta A_U$$

where $\delta A(t)$ is the instantaneous increase in attenuation due to non-linearity and δA_U is the increase which occurs when the disturbing wave is unmodulated. The calculation of δA_U was discussed in Section 5.

6.1. Cross-modulation

The wanted wave is any wave, however weak, traversing the same region of the ionosphere as the disturbing wave. The variation in the attenuation of the wanted wave, due to the disturbing wave, causes the modulation of the latter to be superimposed on the wanted wave. The strength of the wanted wave is also reduced.

Let V_0 be the carrier amplitude of the received signal in the absence of the disturbing wave. When the disturbing wave is present V_0 is decreased by $\delta A(t)$ decibels to a new amplitude $V(t)$. We therefore have

$$20\log_{10}\frac{V_0}{V(t)} = 8{\cdot}68\log\frac{V_0}{V(t)} = \delta A(t)$$

Provided $\delta A(t)$ is less than 2 dB we may write

$$V(t) = V_0\exp\left[-\frac{\delta A(t)}{8{\cdot}68}\right] \simeq V_0\left[1-\frac{\delta A(t)}{8{\cdot}68}\right]$$

$$= V_0\left[1-\frac{f(t)'\,\delta A_U'}{8{\cdot}68}\right]$$

where $\delta A_U'$ is the increase in the attenuation of the wanted wave caused by the disturbing wave when the latter is unmodulated.**

Substitution of an expression for $f(t)'$ similar to Equation (14) then leads to the following results:

(i) The strength of the wanted wave is reduced by

$$\delta A_F' = \delta A_U'\left(1+\frac{(M')^2}{2}\right)\text{dB} \qquad (18)$$

where M' is the modulation index of the disturbing wave.

(ii) The modulation of the disturbing wave is superimposed on the wanted transmission with a level given by

$$L_1' = 20\log_{10}(M'\cos\phi_1'\,\delta A_U') + \delta A_F' - 12{\cdot}8\text{ dB} \qquad (19)$$

(iii) The second harmonic of the disturbing wave is also superimposed with level

$$L_2' = 20\log_{10}\left((M')^2\cos\phi_2\,\delta A_U'\right) + \delta A_F' - 24{\cdot}8\text{ dB} \qquad (20)$$

---

* The linear relationship between the attenuation increase and the e.r.p., referred to in Section 5, may be demonstrated by making the substitution $\hat{E}_U(z) = a(z)\hat{E}_U(0)$ in Equation (16), where $a(z)$ is a function which describes how $\hat{E}_U(z)$ varies along the path of the wanted wave. The attenuation increase is then seen to be proportional to the square of $\hat{E}_U(0)$, the field strength at the point where the wave enters the ionosphere; $[\hat{E}_U(0)]^2$ is proportional to the e.r.p.

** Certain quantities in this section are primed to make the notation consistent with that of Section 6.3.

In Equations (19) and (20) the reference level (0 dB) corresponds to 100% cross-modulation.

6.2. Self-distortion

The wanted wave and the disturbing wave are one and the same. At the modulation peaks the increase in attenuation due to non-linearity is greater than it is when the wave is unmodulated; at the modulation troughs it is less. Non-linearity therefore reduces the peak-to-trough amplitude of the received signal (assumed to be derived from a detector in the absence of automatic gain control) and also generates harmonic distortion.*

In the absence of non-linearity the received signal would be $V_0(1 + M\cos\omega_m t)$; because of non-linearity it is $V(t)$. We therefore have

$$20 \log_{10} \left[\frac{V_0(1 + M\cos\omega_m t)}{V(t)}\right] = \delta A(t)$$

Hence

$$V(t) \simeq V_0\,(1 + M\cos\omega_m t)\left[1 - \frac{\delta A(t)}{8{\cdot}68}\right] \qquad (21)$$

provided $\delta A(t)$ is less than 2 dB. Now $\delta A(t)$ is equal to $f(t)\ \delta A_U$, where $f(t)$ is given by Equation (14). If this value of $\delta A(t)$ is substituted in Equation (21) and the latter expanded, the following approximate results are obtained, provided δA_U does not greatly exceed 0·5 dB.

(i) The carrier amplitude of the received signal is reduced by

$$\delta A_U \left[1 + \frac{M^2}{2} + M^2\cos^2\phi_1\right] \text{ dB}$$

(ii) The fundamental component of the received signal is reduced by

$$\delta A_F = \delta A_U \left[1 + 2\cos^2\phi_1 + \frac{M^2}{4}\,(2 + \cos^2\phi_2)\right] \text{ dB} \qquad (22)$$

(iii) The second-harmonic level (relative to the fundamental) is

$$L_2 = 20 \log_{10}\{M[\cos^2\phi_1 + \cos\phi_1\cos\phi_2\cos(\phi_1 - \phi_2) + \tfrac{1}{4}\cos^2\phi_2]^{\frac{1}{2}}\,\delta A_U\} + \delta A_F - 18{\cdot}8 \text{ dB} \qquad (23)$$

where δA_F is the reduction of the fundamental, given by Equation (22).

(iv) The third-harmonic level is

$$L_3 = 20 \log_{10}(M^2\cos\phi_2\ \delta A_U) + \delta A_F - 30{\cdot}8 \text{ dB} \qquad (24)$$

At very low modulation frequencies the fundamental is reduced by $3(1 + M^2/4)\ \delta A_U$ dB; this is more than three times the reduction of the unmodulated wave, δA_U. At high modulation frequencies, however, the reduction is only slightly greater than δA_U. The reduction lies about midway between the two limiting values when $\omega_m = G\nu_0$ i.e. when the modulation frequency is about 130 Hz. The harmonic distortion is greatest at low modulation frequencies and decreases at 6 dB per octave at high modulation frequencies.

The most important effect is the reduction of the fundamental component of the received signal because this is what determines the reduction in receiver signal-to-noise ratio. It is mainly due to the increase in the attenuation of the signal as a whole rather than to a reduction of the modulation depth. At high modulation frequencies the carrier and sidebands are reduced by exactly the same amount and there is no change in the modulation depth. Only at low modulation frequencies is the modulation depth reduced, but this reduction is small compared with reduction of the carrier frequency. Recent experiments[13] with high power l.f. transmitters have confirmed that demodulation is unimportant, as no significant difference between the percentage modulation of transmitted and received waves was observed.

6.3. Interaction between Two Strong Waves

We consider now the situation which arises when the wanted wave and the disturbing wave are of comparable strength; this could occur, for example, when a station transmits simultaneously on two frequencies.

We make the initial assumption that the work done on an electron by two waves of different frequency is equal to the sum of the work which each wave would do separately. If the arguments contained in the previous sections are followed on this basis we arrive at the result

$$\delta A(t) = \delta A_U f(t) + \delta A'_U f(t)' \qquad (25)$$

---

* The theory contained in this section is similar to that given by King[12] but differs in detail because King's expressions are based on a cross-modulation formula of Huxley and Ratcliffe[1] which neglects the reduction of the carrier amplitude of the wanted wave caused by the disturbing wave.

where δA is the total increase in attenuation
$\delta A_U f(t)$ is the increase caused by the wanted wave on its own
$\delta A'_U f(t)'$ is the increase caused by the unwanted wave on its own*

The received signal $V(t)$ is again given by Equation (21), but $\delta A(t)$ now has the value given by Equation (25). Expansion of Equation (21) then shows that the fundamental component of the wanted modulation is reduced by $\delta A_F + \delta A'_F$, given by Equations (18) and (22). The harmonic distortion levels of the wanted transmission are given by Equations (23) and (24) but with δA_F replaced by $\delta A_F + \delta A'_F$.

The modulation of the unwanted transmission is superimposed on the wanted transmission, together with its second harmonic. The cross-modulation levels derived from Equation (21) are:

Fundamental

$$L'_1 = 20 \log_{10}(M' \cos\phi_1 \ \delta A'_U) + \delta A_F + \delta A'_F - 12{\cdot}8 \text{ dB} \tag{26}$$

Second harmonic

$$L'_2 = 20 \log_{10}\left((M')^2 \cos\phi_2 \ \delta A'_U\right) + \delta A_F + \delta A'_F - 24{\cdot}8 \text{ dB} \tag{27}$$

These levels express the absolute modulation of the wanted transmission caused by the unwanted transmission; for example, a level of −20 dB would correspond to 10% modulation of the wanted transmission by the unwanted transmission. The levels also depend to a small extent on the depth of modulation of the *wanted* transmission, but this is a second-order effect; Equations (26) and (27) give the levels occurring when the wanted transmission is unmodulated

Intermodulation frequencies $f_m \pm f'_m$ and $f_m \pm 2f'_m$, where f_m and f'_m are the modulation frequencies of the wanted and unwanted transmissions respectively, are also generated. The intermodulation levels given by Equation (21), relative to the level of the wanted modulation, are:

Frequencies $f_m \pm f'_m$

$$L_{11} = 20 \log_{10}(M' \cos\phi_1 \ \delta A'_U) + \delta A_F + \delta A'_F - 18{\cdot}8 \text{ dB} \tag{28}$$

Frequencies $f_m \pm 2f'_m$

$$L_{12} = 20 \log_{10}\left((M')^2 \cos\phi_2 \ \delta A'_U\right) + \delta A_F + \delta A'_F - 30{\cdot}8 \text{ dB} \tag{29}$$

Radio-frequency intermodulation products, such as the sum and difference of the two carrier frequencies, are not generated because the ionospheric time constant $1/G\nu_0$ is very large compared with the period of one cycle at r.f. No trace of these frequencies was detected in the recent experiments with high-power European transmitters[13] although it was expected that they would be observed.

7. CONCLUSIONS

When a powerful disturbing wave traverses the ionosphere the energy balance of the free electrons is upset and their collision frequency increased. As a result the attenuation suffered by this and other waves traversing the same region of the ionosphere may be increased. The attenuation increase is proportional to the power of the disturbing transmitter and inversely proportional to its frequency unless it excites the extraordinary wave strongly, in which case the attenuation is accentuated near the gyromagnetic frequency.

If the disturbing wave is amplitude modulated the attenuation will vary over the modulation cycle and the modulation will be impressed on all waves traversing regions of the ionosphere where significant attenuation changes occur. Medium-frequency waves are modulated in this manner in the E-layer but not in the D-layer. The cross-modulation is greatest at low audio frequencies because the changes in collision frequency do not occur instantly but are subject to a time constant of about 1·25 milliseconds.

If the disturbing wave itself is being received the attenuation increase will result in a non-linear relationship between the transmitted and received powers and the modulation waveform will be distorted. Ionospheric non-linearity may also give rise to audio-frequency intermodulation products if two powerful transmitters illuminate a common region of the ionosphere but radio-frequency intermodulation products will not be generated because of the large ionospheric time constant.

In view of the importance of these effects where high-power transmitters are concerned, expressions have been derived and are given in this report for:

(a) the reduction and harmonic distortion of the received signal, caused by non-linearity

(b) the cross-modulation suffered by all waves traversing the region of the ionosphere illuminated by a high-power transmitter

(c) the cross-modulation and intermodulation which arises when two high-power transmitters are situated in the same locality.

* Throughout this section primed symbols refer to the unwanted wave.

8. REFERENCES

1. BAILEY, V.A. 1937. On some effects caused in the ionosphere by electric waves. *Phil. Mag.* 1937, **23**, 157, pp. 929-960. 1938, **26**, 176, pp. 425-453.

2. BAILEY, V.A. 1963. Some non-linear phenomena in the ionosphere. (Conference on non-linear processes in the ionosphere, Boulder, Colorado, December 16-17, 1963), *NBS Technical Note* No. 211, April 1964, **3**, pp. 1-60.

3. CUTOLO, M. 1963. On some resonance phenomena near the gyrofrequency obtained during the propagation of a radio wave in plasma. ibid **2**, p.1-69.

4. HIBBERD, F.H. 1963. Wave interaction at oblique incidence, near the gyro frequency. ibid **4**, p.45-64.

5. HUXLEY, L.G.H. *and* RATCLIFFE, J.A. 1949. A survey of ionospheric cross-modulation. *Proc. Instn. elect. Engrs.*, 1949, **96**, III, p.433-40.

6. GINZBURG, V.L. *and* GUREVICH, A.V. 1960. Non-linear phenomena in a plasma located in an alternating electromagnetic field. *Soviet Phys. Usp.*, 1960, **3**, pp. 115-146, 1960, **3**, pp. 175-195 in English.

7. BAILEY, V.A. 1937. The motion of electrons in a gas in the presence of variable electric fields and a constant magnetic field. *Phil. Mag.*, 1937, **23**, 156, pp. 774-791.

8. RATCLIFFE, J.A. 1959. The magneto-ionic theory and its application to the ionosphere. Cambridge, Cambridge University Press, 1959.

9. DEEKS, D.G. 1965. D-region distributions in middle latitudes deduced from the reflexion of long radio waves. *Proc. Royal Soc.*, *A*, 1966, **291**, 1426, pp. 413-437.

10. HALL, J.E. *and* FOOKS, J. 1965. The electron distribution in the quiet D-region derived from rocket measurements of low-frequency propagation. *Planet. Space. Sci.*, 1965, **13**, p. 1013-1030.

11. EBERT, W. 1962. Ionospheric propagation on long and medium waves. *EBU Document*, 1962, TECH **3081-E**.

12. KING, J.W. 1958. Ionospheric self-demodulation and self-distortion of radio waves, *J. atmos. terr. Phys.*, 1959, **14**, pp. 41-49.

13. HABERKANT, E. *and* VOGT, K. 1966. Measurement of ionospheric cross-modulation affecting two high-power l.f. broadcast transmitters. *E.B.U. Rev.*, 1966, **98A**, pp. 154-161.

Printed by BBC Research Department, Kingswood Warren, Tadworth, Surrey